



Sampling Image Segmentations for Uncertainty Quantification

Matthieu Lê, Jan Unkelbach, Nicholas Ayache, Hervé Delingette

► To cite this version:

Matthieu Lê, Jan Unkelbach, Nicholas Ayache, Hervé Delingette. Sampling Image Segmentations for Uncertainty Quantification. Medical Image Analysis, 2016, 34, pp.42-51. 10.1016/j.media.2016.04.005 . hal-01304646

HAL Id: hal-01304646

<https://hal.science/hal-01304646>

Submitted on 20 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sampling Image Segmentations for Uncertainty Quantification

Matthieu Lê^a, Jan Unkelbach^b, Nicholas Ayache^a, Hervé Delingette^a

^a*Asclepios Project, Inria Sophia Antipolis, France.*

^b*Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.*

Abstract

In this paper, we introduce a method to automatically produce plausible image segmentation samples from a single expert segmentation. A probability distribution of image segmentation boundaries is defined as a Gaussian process, which leads to segmentations which are spatially coherent and consistent with the presence of salient borders in the image. The proposed approach is computationally efficient, and generates visually plausible samples. The variability between the samples is mainly governed by a parameter which may be correlated with a simple Dice's coefficient, or easily set by the user from the definition of probable regions of interest. The method is extended to the case of several neighboring structures, but also to account for under or over segmentation, and the presence of excluded regions. We also detail a method to sample segmentations with more general non-stationary covariance functions which relies on supervoxels. Furthermore, we compare the generated segmentation samples with several manual clinical segmentations of a brain tumor. Finally, we show how this approach can have useful applications in the field of uncertainty quantification, and an illustration is provided in radiotherapy planning, where segmentation sampling is applied to both the clinical target volume and the organs at risk.

Keywords: Segmentation, uncertainty, Gaussian process, radiotherapy planning, brain tumor

1. Introduction

Medical image segmentation is a key technology for many medical applications, like computer aided diagnosis or therapy planning and guidance. Due to its ill-posed nature, the quantification of segmentation uncertainty is crucial to assess the overall performance of other applications. In radiotherapy planning for instance, it is important to estimate the impact of uncertainty in the delin-

Email address: `matthieu.le@inria.fr` (Matthieu Lê)

eation of the gross tumor volume and the organs at risk on the dose delivered to the patient.

A straightforward way to assess this impact is to perform *Image Segmentation Sampling* (ISS), which consists of gathering several plausible segmentations of the same structure, and estimate the variability of the output due to the variability of the segmentations. In this paper, a segmentation is said plausible if it is spatially coherent and consistent with the presence of salient borders in the image. For computer generated segmentations, ISS could simply be obtained by varying the parameters or initial values of the algorithm producing the segmentations. However, in many cases, parameters of the algorithms cannot be modified, and segmentations are partially edited by a user. For manual or semi-manual segmentations, it is possible to estimate the inter-expert variability on a few cases but it usually cannot be applied on large databases due to the amount of resources required.

This is why it is important to automate the generation of plausible segmentations, which are "similar to" a given segmentation of a region of interest (ROI). This is the objective of this paper which, to the best of our knowledge, has not been tackled before. It is naturally connected to several prior work in the field of image segmentation. Indeed, a large amount of generative models have been developed to compute segmentations based on a single image. For instance, deformable models (McInerney and Terzopoulos (1996)) or Markov Random Fields (MRF) (Wang et al. (2013)) have been extensively studied in order to combine a data term and a smoothness prior term to compute realistic segmentations. These models are usually used to infer the maximum a posteriori segmentation, and rarely to sample from the model which can be computationally expensive (Niethammer et al. (2015)).

(Chang and Fisher III (2011); Fan et al. (2007)) have proposed segmentation approaches based on Markov Chain Monte Carlo where parameter sampling leads to an estimation of the posterior probability of obtaining a segmentation given an image. In those approaches however, the algorithm defines the likelihood and prior functions and then estimates the most probable (or the expected) segmentation whereas in ISS the objective is to sample directly from the posterior distribution, knowing only its mean or mode.

Other methods (De Bruijne and Nielsen (2004); Petersen et al. (2010)) rely on sampling in order to find the most probable segmentation. It is particularly useful when the object of interest is not clearly visible or partially occluded. In this case, segmentation methods using sampling algorithms have been developed in order to introduce prior knowledge on the shape of the expected segmentation. This prior knowledge can for instance result from a point distribution model defined using a database of previously segmented shapes. This is not directly applicable in our case since we do not assume such a database is available. Moreover, it may not necessarily be useful in the case of brain tumors for instance, which do not have a typical shape.

Other related approaches (Pohl et al. (2007); Sabuncu et al. (2010); Warfield et al. (2004)) aim at producing a consensus segmentation given several expert segmentations, or several atlas segmentations. They define probabilities of hav-

ing a given segmentation based on a reference one, and their generative nature makes them suitable for ISS. Typical examples are the STAPLE algorithm (Warfield et al. (2004)), the log-odds maps (Pohl et al. (2007)) and their refinement (Sabuncu et al. (2010)). However, as shown in section 2, the segmentations generated from a single expert segmentation lack plausibility, and the spatial regularity of the contours cannot be finely controlled.

In this paper, a novel framework is introduced to sample segmentations automatically leading to visually plausible delineations. More precisely, the proposed approach incorporates knowledge about image saliency of the ROI such that the sampled contours’ variability may be greater at poorly contrasted regions, and smaller near sharp image gradients. Furthermore, the proposed approach is mathematically well grounded, and enforces the spatial smoothness of the contours, because it relies on Gaussian processes defined on implicit contours. Finally, segmentation sampling can be performed efficiently even on large medical images thanks to an algorithm using the structure of the image and the covariance matrix. Variability in the samples is easily controlled by a single scalar. We also exhibit an application of the method to radiotherapy dose planning.

This paper is an extension of (Lê et al. (2015b)). We provide more insights on the algorithms which can be used for efficient sampling. We also introduce a new section detailing possible extensions of the method. We detail how it can be applied in the case of several neighboring ROIs, in the case of under or over segmentation, and to force structures to be included or excluded from the samples. Moreover, we introduce a method to sample segmentations using general covariance matrices which is based on supervoxels. We also compare our method to several segmentations from different clinicians to evaluate the plausibility of the generated samples. Finally, the application to radiotherapy planning has been extended by taking into account the impact of adjacent organs at risk.

2. Existing Generative Models of Segmentations

This section reviews relevant generative models of segmentations proposed in the literature. Results are illustrated on a synthetic image (Figure 1) for which the structure border is surrounded by regions of low and high contrast.

The probabilistic atlases (Pohl et al. (2007)) derived from log-odds of signed distance functions assume that voxels are independently distributed with a Bernoulli probability density function of parameter b whose value depends on the distance to the structure border. The STAPLE algorithm (Warfield et al. (2004)) is a region formulation for producing consensus segmentations. Given a binary segmentation T , and expert sensitivity p and specificity q , the algorithm is associated with a generative model for which a segmentation D can be sampled knowing T as a Markov Random Field with the likelihood term

$$P(D_i = 1) = pP(T_i = 1) + (1 - q)P(T_i = 0) \quad (1)$$

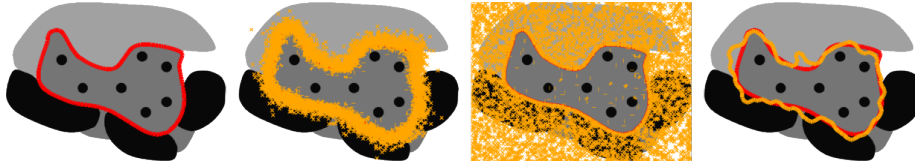


Figure 1: From left to right: synthetic image with region of interest outlined in red; segmentation sampling based on log-odds; segmentation sampling based on STAPLE without ICM steps ($p = 97\%$ and $q = 97\%$); segmentation sampling based on STAPLE with ICM steps ($p = 97\%$ and $q = 97\%$). The ground truth is outlined in red, the samples are outlined in orange.

and a prior accounting for local spatial coherence. Segmentations are generated by sampling independently the Bernoulli distribution at each voxel followed by a number of Iterated Conditional Modes (ICM) relaxation steps. Various ISS results are obtained in Figure 1 for the log-odds and STAPLE generative models with specified parameters.

In all cases, the produced segmentations are not realistic for two reasons. First, the variability of the segmentation does not account for the intensity in the image such that borders with strong gradients are equally variable as borders with weak gradient. This is counter intuitive as the basic hypothesis of image segmentation is that changes of intensity are correlated with changes of labels. Second, borders of the segmented structures are unrealistic mainly due to their lack of geometric regularity (high frequency wobbling in Figure 1 (Right)). While anatomical or pathological structure borders are not necessarily smooth (e.g. highly diffuse tumors), the generated samples show irregular generated contours in the presence of regular visible contours in the image, which is not plausible.

3. GPSSI

3.1. Definition

We propose a generative model of image segmentation which overcomes the two limitations of the presented previous approaches. First, sampled segmentations do take into account the image intensity by replacing the signed distance with signed geodesic distance. Second, spatial consistency of the sampled segmentations is enforced by describing a probabilistic segmentation with a Gaussian process with a squared exponential covariance, which allows to easily control the spatial coherence of the segmentation. The geodesic distance makes voxels far away from the mean segmentation if they are separated from it by high gradient intensity regions. Therefore, a random perturbation on the mean segmentation is unlikely to reach those voxels with high contrast, and more likely to affect voxels with low geodesic distance, i.e. voxels neighboring the mean segmentation with similar intensity values.

A novel probabilistic framework of image segmentation is introduced by defining a level set function via a Gaussian process (GP). We name the method

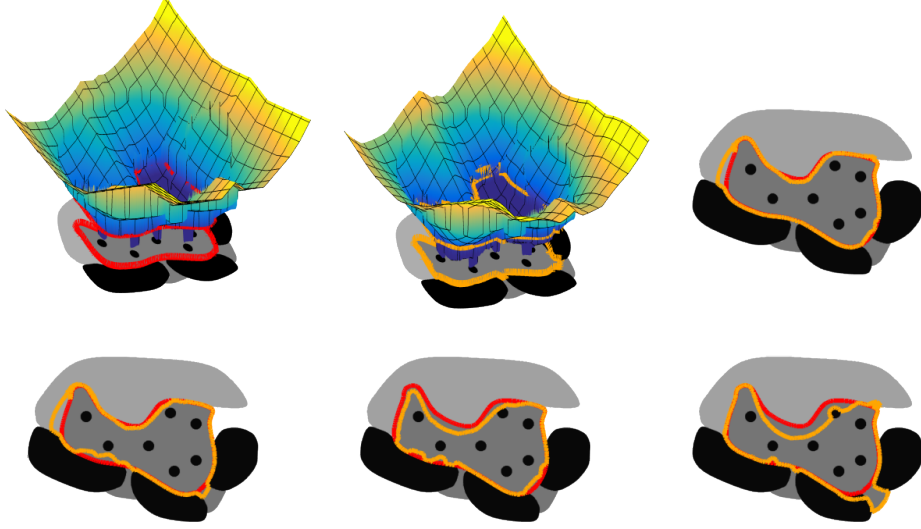


Figure 2: (Top Left) Mean of the GP μ ; (Top Middle) Sample of the level set function $\varphi(\mathbf{a})$ drawn from $\mathcal{GP}(\mu, \Sigma)$; (Others) GPSSI samples. The ground truth is outlined in red, the GPSSI samples are outlined in orange.

GPSSI for Gaussian Process for Sampling Segmentations of Images. The mean of the GP is given by a signed geodesic distance, and its covariance is defined with a squared exponential driven by the Euclidean distance between voxels. Gaussian process implicit surfaces have been introduced previously by Williams and Fitzgibbon (2007) as a generalization of thin plate splines, and used recently by Gerardo-Castro et al. (2013) for surface reconstruction. However, our approach combining geodesic and Euclidean distance functions for the mean and covariance is original, and specifically suited to represent probabilistic image segmentations.

3.2. Geodesic Distance Map

Signed geodesic distance maps are computed as

$$\mathcal{G}(\mathbf{a}) = \min_{\Gamma \in \mathcal{P}_{\text{seg}, \mathbf{a}}} \int_0^1 \sqrt{\|\Gamma'(s)\|^2 + \gamma^2 \left(\nabla \mathcal{I}(\Gamma(s)) \cdot \frac{\Gamma'(s)}{\|\Gamma'(s)\|} \right)^2} ds \quad (2)$$

where \mathcal{I} is the input image, $\mathcal{P}_{\text{seg}, \mathbf{a}}$ is the set of all paths between the voxel \mathbf{a} and the segmentation, and Γ one such path, parametrized by $s \in [0, 1]$, with spatial derivative $\Gamma'(s) = \partial \Gamma(s) / \partial s$. The parameter γ sets the trade-off between Euclidean distance ($\gamma = 0$) and gradient information. Its implementation is based on a fast grid sweeping method as proposed by Toivanen (1996) and used in Criminisi et al. (2008). The gradient is computed with a Gaussian kernel convolution controlled by the parameter h . The signed geodesic distance is set negative inside the segmentation, and positive outside.

3.3. GPSSI

GP are a generalization of multivariate Gaussian distributions, and provide a framework to define probability distributions over functions. GP are widely used in machine learning for solving inference problems (Williams and Rasmussen (2006)) over spatially correlated datasets. In this paper, it is the generative nature of GP which is of interest, since they naturally produce spatially smooth samples.

In GPSSI, a segmentation over a set Ω is defined via a level set function $\varphi(\mathbf{a})$, $\mathbf{a} \in \Omega$ such that its zero level set corresponds to the boundary of the ROI. Smoothness in the level set function φ translates into the smoothness of the boundary $\mathcal{B}_\varphi = \{\mathbf{a} \in \Omega \mid \varphi(\mathbf{a}) = 0\}$. A GP is fully defined by its mean and covariance functions: its mean value is set to the signed geodesic distance $\mu(\mathbf{a}) = \mathcal{G}(\mathbf{a})$ while its covariance is chosen as the squared exponential function,

$$\forall \mathbf{a}, \mathbf{b} \in \Omega, \quad \Sigma(\mathbf{a}, \mathbf{b}) = \omega_0 \exp\left(-\frac{\|\mathbf{a} - \mathbf{b}\|^2}{\omega_1^2}\right) \quad (3)$$

This choice of covariance enforces the smoothness of the segmentation, with parameter ω_1 characterizing the typical correlation length between two voxels while ω_0 controls the amount of variability of the level set function.

3.4. Efficient Sampling

Sampling of a GP is simply performed through the factorization of the covariance matrix at sample points. More precisely, let $\Omega_M = \{\mathbf{a}_i\}$, $i = 1 \dots M$ be the set of M discrete points \mathbf{a}_i where the level set function φ is defined. Typically, Ω_M may be the set of all voxel centers in the image. The covariance matrix $\Sigma_{ij}^{MM} = \omega_0 \exp(-\|\mathbf{a}_i - \mathbf{a}_j\|^2/\omega_1^2)$ at sampled points is of size $M \times M$. To sample from a GP $\mathcal{GP}(\mu, \Sigma)$, a factorization of the covariance matrix $\Sigma^{MM} = LL^\top$ is required, such that given normally distributed variables $u \sim \mathcal{N}(0, \mathbb{I})$, GPSSI are simply computed as the zero crossing of $\mu + L(\omega_0, \omega_1)u \sim \mathcal{GP}(\mu, \Sigma)$.

A classical issue with GP sampling is that the factorization of Σ^{MM} becomes computationally expensive and ill-conditioned for large values of M . Since in practice $M \approx 10^7$, a regular matrix factorization - usually a Cholesky decomposition - in $\mathcal{O}(M^3)$ is computationally prohibitive. However, several methods exist to efficiently sample high dimensional GPs when the sample points form a regular grid and the covariance matrix is stationary (i.e. invariant by translation):

1. **Periodic boundary conditions** (Dietrich and Newsam (1997); Kozintsev (1999); Kozintsev and Kedem (2000)). Assuming periodic boundary conditions on the image, Σ^{MM} is a Block Circulant with Circulant Blocks (BCCB) matrix such that each row of Σ^{MM} is a periodic shift of the first row of Σ^{MM} , $C \in \mathbb{R}^M$. C can be seen as an image of M voxels, whose voxel value is the evaluation of the squared exponential covariance for every shift present in the image. Theoretical results on the BCCB matrix spectral decomposition give us

$$\Sigma^{MM} = F^{-1} \text{diag}(FC)F \quad (4)$$

where the complex matrix F is the $M \times M$ discrete Fourier transform matrix (Kozintsev (1999)). Hence, the eigenvalues of Σ^{MM} are the discrete Fourier transform of C . As such, if $u_1, u_2 \sim \mathcal{N}(0, \mathbb{I})$ i.i.d, then the real and imaginary parts of

$$F \sqrt{\text{diag}(FC)}(u_1 + iu_2) \quad (5)$$

are two independent samples from the GP (Kozintsev (1999)). This can be efficiently computed using the Fast Fourier Transform without storing F . Samples can then be generated in $\mathcal{O}(M \log(M))$ (Figure 3 Top Left).

2. **Circulant embedding of the Toeplitz matrix** (Dietrich and Newsam (1997); Kozintsev (1999); Kozintsev and Kedem (2000)). If periodic boundary conditions are not applied, Σ^{MM} is a block Toeplitz with Toeplitz blocks matrix. This matrix can be embedded in a larger Σ^{2M2M} BCCB matrix. As previously, samples can then be drawn in $\mathcal{O}(M \log(M))$ (Figure 3 Top Left).
3. **Kronecker product** (Gilboa et al. (2015); Lorenzi et al. (2015); Saatçi (2012)). Another approach is to note that the chosen covariance function is separable. In the 3-dimensional case,

$$\forall \mathbf{a}, \mathbf{b} \in \Omega,$$

$$\Sigma^{MM}(\mathbf{a}, \mathbf{b}) = \omega_0 \exp\left(-\frac{(\mathbf{a}_x - \mathbf{b}_x)^2}{\omega_1^2}\right) \exp\left(-\frac{(\mathbf{a}_y - \mathbf{b}_y)^2}{\omega_1^2}\right) \exp\left(-\frac{(\mathbf{a}_z - \mathbf{b}_z)^2}{\omega_1^2}\right) \quad (6)$$

Then,

$$\Sigma^{MM} = \Sigma^{M_x M_x} \otimes \Sigma^{M_y M_y} \otimes \Sigma^{M_z M_z} \quad (7)$$

where $\Sigma^{M_x M_x} \in \mathbb{R}^{M_x \times M_x}$, $\Sigma^{M_y M_y} \in \mathbb{R}^{M_y \times M_y}$, and $\Sigma^{M_z M_z} \in \mathbb{R}^{M_z \times M_z}$ are the covariance matrices along the 3 dimensions, and \otimes denotes the Kronecker product. This way, a Cholesky decomposition of Σ^{MM} can be obtained through the individual factorization of $\Sigma^{M_x M_x}$, $\Sigma^{M_y M_y}$, and $\Sigma^{M_z M_z}$. Samples can then be drawn in $\mathcal{O}(M_x \log(M_x))$, assuming $M_x > M_y$, and $M_x > M_z$ (Figure 3 Top Left). Moreover, leveraging matrix/vector products and properties of the Kronecker product, the full factorization of Σ^{MM} needs not be stored (Gilboa et al. (2015)).

We first check the computation times (Figure 3 top left). We can see that using the Kronecker product takes more time than the Circulant embedding of the Toeplitz matrix, which in turn is more time consuming than assuming periodic boundary condition on the image. However, we noticed that the Kronecker product is more stable for large ω_1 . As such, we chose this method in the rest of the paper since the size of the used images was not computationally prohibitive.

Examples of samples using the three methods can be seen on Figure 3 (top). We further check the samples against the ground truth. We compute the sample correlation between the pixel on the top left and the pixels of the first row using 5000 samples. The computed correlation is close to the theoretical correlation $f_1(a) = \omega_0 \exp(-a^2/\omega_1^2)$. We also plot the histogram of the 5000 samples of a random pixel and retrieve the density $f_2(x) = 1/\sqrt{2\pi\omega_0} \exp(-x^2/(2\omega_0))$.

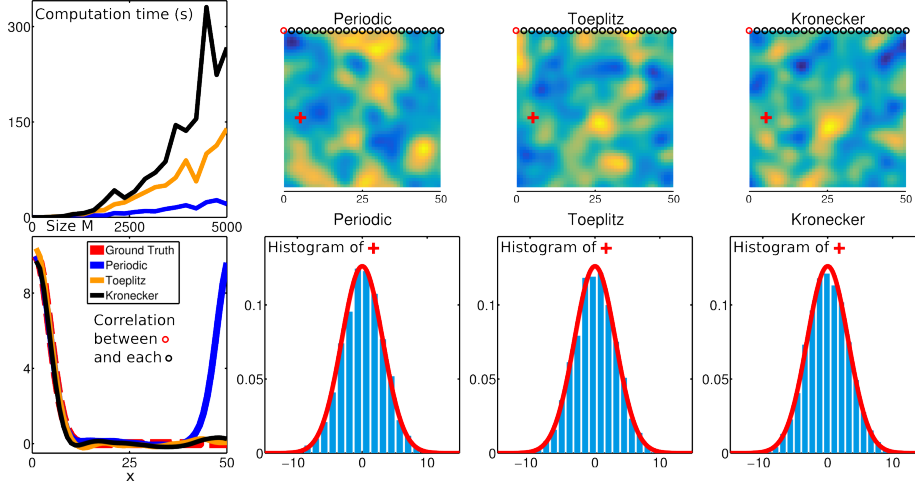


Figure 3: (Top row, from left to right) Computation times (seconds) for the sampling of 3 $M \times M$ Gaussian processes with $\omega_0 = 10$ and $\omega_1 = 5$, function of the size M . Example of a 50×50 sample using periodic boundary conditions on the image, circulant embedding of the Toeplitz matrix, and the Kronecker decomposition. (Bottom row, from left to right) Sample correlation between the point at the top left corner (red circle) and the points of the first row (black circles) for 5000 Gaussian process samples of size 50×50 . The red line is the expected correlation $f_1(a) = \omega_0 \exp(-a^2/\omega_1^2)$. Histograms of 5000 realizations of a point (red cross) of a 50×50 Gaussian process using periodic boundary conditions on the image, circulant embedding of the Toeplitz matrix, and the Kronecker decomposition. The red line represents the expected density $f_2(x) = 1/\sqrt{2\pi\omega_0} \exp(-x^2/(2\omega_0))$.

4. Parameter Settings

In the proposed approach, segmentation sampling depends on the scale h of the gradient operator, the parameter γ of the geodesic map, and the parameters ω_0 and ω_1 of the covariance function. The parameter h depends on the level of noise in the image (typically chosen as 1 voxel size) whereas γ controls the importance of the geodesic term. In our experiments, we set $\gamma = 100/\mathbb{E}(\mathcal{I})$, where $\mathbb{E}(\mathcal{I})$ is the mean of 5th to the 95th percentiles of the image intensity.

The parameter ω_1 controls the smoothness scale of the structure, and is chosen as the radius of the equivalent sphere. In 3D, given the volume V of the ROI, we set $\omega_1 = (3/(4\pi)V)^{1/3}$.

The parameter ω_0 controls the variability around the mean shape: the greater ω_0 , the greater the variability. Such variability may be practically quan-

tified for instance in terms of mean inter-sample Dice’s coefficient between every pairs of expert segmentations. In this case, it is easy to find ω_0 corresponding to a given mean inter-sample Dice’s coefficient (see Figure 4 middle right). This approach offers an intuitive way to semi-automatically set the parameter ω_0 . Another way of setting the parameter ω_0 is to relate it to the mean Dice’s coefficient between the samples and the input segmentation (see Figure 4 middle left). Instead of Dice’s coefficient, one could also use quantiles of histograms of symmetric distances between contours.

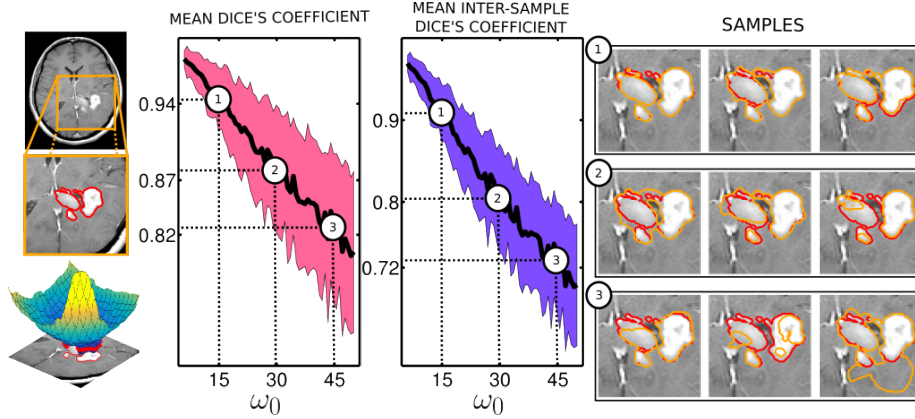


Figure 4: (Left) Segmentation of brain tumor active rim from T1 MR image with Gadolinium contrast agent; (Middle Left) Relationship between the parameter ω_0 and the mean Dice’s coefficient between 40 samples and the clinical segmentation; (Middle Right) Relationship between the parameter ω_0 and the mean inter-sample Dice’s coefficient using 40 samples; (Right) Samples for different ω_0 . The clinician segmentation is outlined in red, the GPSSI samples are outlined in orange.

Another approach is to let a user define a confidence region where the segmentations should lie. Let the user choose the two tightest isocontours $\pm D$ of the geodesic distance which enclose this confidence region. For a voxel \mathbf{a} on the defined isocontours, we have $\varphi(\mathbf{a}) \sim \mathcal{N}(\mu(\mathbf{a}) = \pm D, \sigma = \sqrt{\omega_0})$. We can then define ω_0 based on the 95% confidence interval for the normal distribution that the sample will not be negative (i.e. included in the segmentation),

$$D = 2\sigma = \sqrt{\omega_0} \implies \omega_0 = \left(\frac{D}{2}\right)^2 \quad (8)$$

Figure 5 shows an example where we set ω_0 such that the samples most probably lie in a region delineated by the isocontours $\mu(\mathbf{a}) = \pm 45$.

5. Segmentation Sampling

Samples of the 2D synthetic segmentation case can be seen on Figure 2 with $\omega_0 = 506$ corresponding to a mean inter-sample Dice’s coefficient of 80%.

Samples are coherent with the visible image boundary since most samples do not include highly contrasted (black) regions of the image but instead invade low contrast regions of the image.

Segmentation sampling was also performed on a 3D T1 post contrast MRI (T1Gd MRI) where the proliferative part (active rim) of a grade IV glioma was segmented by an expert (Figure 4 left). The strong correlation between the covariance parameter ω_0 and the mean inter-sample Dice's coefficient as well as the mean Dice's coefficient against the clinician segmentation was computed after generating 40 samples for each value of ω_0 (Figure 4 right). Thus the user may easily choose ω_0 as a function of the desired Dice's coefficient.

Note that the likelihood of samples generated from $\mathcal{GP}(\mu, \Sigma)$ is not very informative as it is computed over the whole image and not just the generated contour.

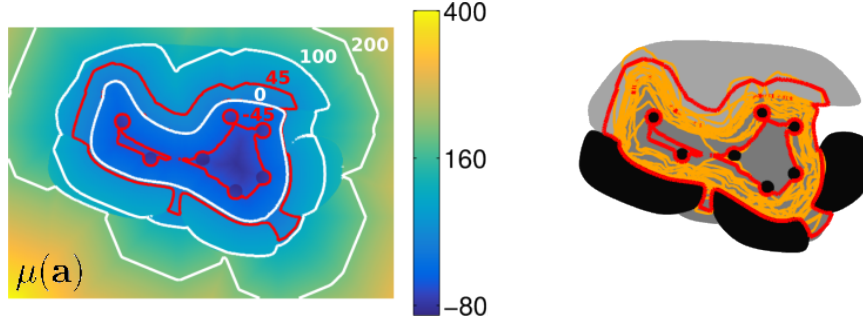


Figure 5: (Left) Signed geodesic distance $\mu(\mathbf{a})$ of the ROI with isocontours -45, 0, 45, 100, 200. (Right) One can check that the samples most probably lie in the region delineated by the isocontours $\mu(\mathbf{a}) = \pm 45$ (in red). The sampled contours are in orange.

6. GPSSI Extensions

6.1. Several Neighboring Structures

The presented method can directly be used to sample several ROIs. However, when the structures are close to each other, the sampled structures may overlap. In such a case, it is important to sample them jointly to define non-overlapping segmentations. Based on Zhao et al. (1996), the different ROIs can be defined using different level set functions. For two structures, one can define two Gaussian processes $\varphi_1 \sim \mathcal{GP}(\mu_1, \Sigma_1)$ and $\varphi_2 \sim \mathcal{GP}(\mu_2, \Sigma_2)$, where μ_1 and μ_2 are the signed geodesic distances from the two considered ROIs, and Σ_1 and Σ_2 are the corresponding covariance matrices. As such, non-overlapping samples of the two ROIs \mathcal{S}_1 and \mathcal{S}_2 can be defined as the ensembles (Figure 6):

$$\mathcal{S}_1 = \{\mathbf{a} \mid \varphi_1(\mathbf{a}) \leq 0 \text{ and } \varphi_1(\mathbf{a}) \leq \varphi_2(\mathbf{a})\} \quad (9)$$

$$\mathcal{S}_2 = \{\mathbf{a} \mid \varphi_2(\mathbf{a}) \leq 0 \text{ and } \varphi_2(\mathbf{a}) < \varphi_1(\mathbf{a})\} \quad (10)$$

Figure 11 shows examples of two structure samples: non-overlapping samples of a glioma and brainstem segmented by a clinician on a T1Gd MRI were computed

using the parameters as defined in the previous section, and a manually set ω_0 for a mean inter-sample Dice's coefficient of 83% for the glioma and 85% for the brainstem. Unlike (Vese and Chan (2002)), one needs as many level set functions as the number of neighboring ROIs for the sampling, because a specific set of parameters ω_0 and ω_1 is associated with each structure. The assignment of labels in the overlapping region is then assigned to the minimum level set function.

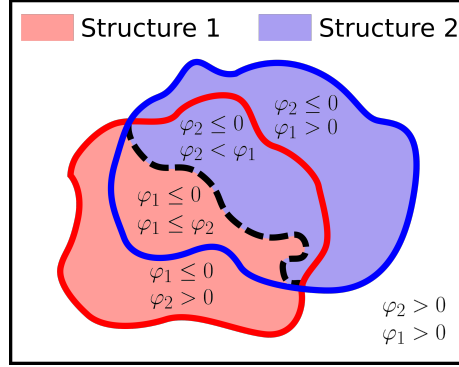


Figure 6: Sampling two non-overlapping structures can be achieved using 2 different level set functions $\varphi_1 \sim \mathcal{GP}(\mu_1, \Sigma_1)$ and $\varphi_2 \sim \mathcal{GP}(\mu_2, \Sigma_2)$.

6.2. Accounting for Under and Over Segmentation

Over or under segmentation can be one of the major differences between the performance of two experts. This effect can be easily incorporated in this framework by adding a variable ϵ such that the level set function $\varphi \sim \mathcal{GP}(\mu + \epsilon, \Sigma(\omega_0, \mu_1))$. If $\epsilon > 0$, the samples will on average result in under segmentation, if $\epsilon < 0$ the samples will on average result in over segmentation. Examples of under and over segmentation can be seen on Figure 7 (Middle) where under segmentation is achieved with $\epsilon = +20$ and over segmentation with $\epsilon = -20$. This parameter can be randomly sampled around 0 when sampling segmentations in order to take under and over segmentation into account.

6.3. Inclusion or Exclusion of Structures

To provide anatomically consistent segmentations, it can be interesting to force specific regions of the image to stay excluded or included from the samples. For instance, it is expected that cerebrospinal fluid or skull regions would not be included in any glioma segmentations. In the GPSSI approach, this could be handled by making the ω_0 parameter spatially varying. However, such a covariance function would not necessarily be stationary nor positive definite. As such, using the fast sampling methods exposed in Section 3.4 would not be possible. Instead we propose to achieve a similar effect by keeping a stationary covariance function, but modifying the mean of the Gaussian process μ . Indeed, we know that $\varphi(\mathbf{a}) \sim \mathcal{N}(\mu(\mathbf{a}), \omega_0)$ for each voxel \mathbf{a} by marginalizing over the

rest of the voxels. We want to force the voxel to stay outside or inside of the samples. This can be done by forcing the variance to a lower value such that $\varphi_1(\mathbf{a}) \sim \mathcal{N}(\mu(\mathbf{a}), \omega_{01}(\mathbf{a}))$ where $\omega_{01}(\mathbf{a}) < \omega_0$. In other words, we are more confident about the labeling of \mathbf{a} . One can note that the probability of \mathbf{a} to be inside the segmentation is

$$P(\varphi_1(\mathbf{a}) \leq 0) = P\left(\varphi(\mathbf{a}) + \mu(\mathbf{a}) \left(\frac{\omega_0}{\omega_{01}(\mathbf{a})} - 1\right) \leq 0\right) \quad (11)$$

where

$$\varphi(\mathbf{a}) + \mu(\mathbf{a}) \left(\frac{\omega_0}{\omega_{01}(\mathbf{a})} - 1\right) \sim \mathcal{N}\left(\mu(\mathbf{a}) \frac{\omega_0}{\omega_{01}(\mathbf{a})}, \omega_0\right) \quad (12)$$

One can then "mimic" this spatially varying confidence that a voxel should be included or excluded from the segmentation by simply normalizing the signed geodesic distance μ with a spatially varying factor. In other words, a Gaussian process with a normalized mean μ has the same voxel-wise marginalized distribution as a Gaussian process with spatially varying ω_0 . Segmentation samples of a glioma on a T2-FLAIR MRI are shown on Figure 7 (Right) where a region around the brainstem and ventricles is forced outside of the samples by setting $\omega_0/\omega_{01} = 2$.

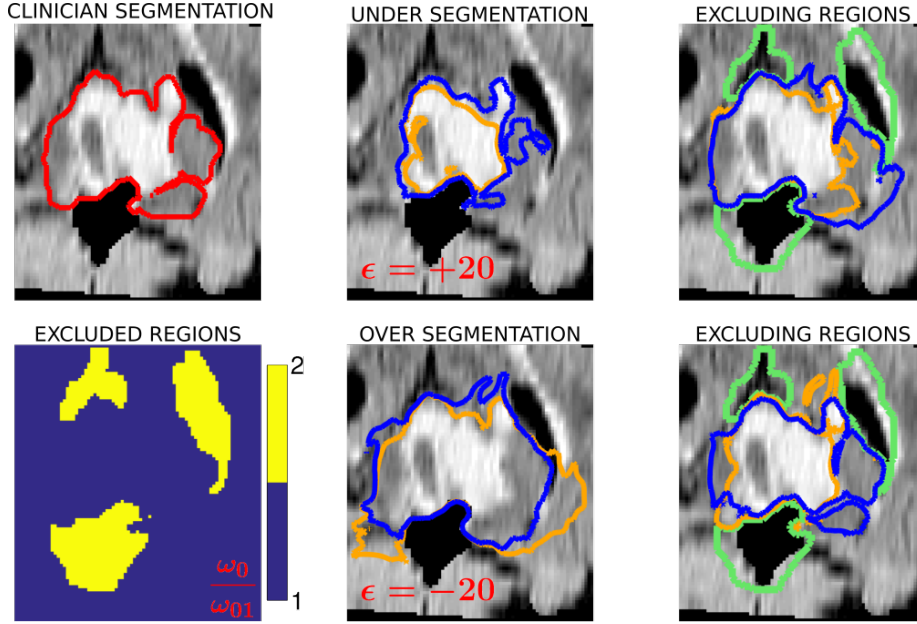


Figure 7: From top to bottom, left to right: clinician segmentation of a glioma on a T2-FLAIR MRI; factor ω_0/ω_{01} used to normalize the signed geodesic distance μ to exclude the region in yellow; under segmentation samples (in orange and blue) using $\epsilon = +20$; over segmentation samples (in orange and blue) using $\epsilon = -20$; samples (in orange and blue) where regions around the brainstem and ventricles (in green) were forced to be excluded by multiplying the signed geodesic distance μ by the spatially varying factor ω_0/ω_{01} .

6.4. General Covariance Matrix Using Supervoxels

In so far, we have detailed a method to efficiently compute segmentation samples using the squared exponential covariance function. A stationary covariance function was necessary in order to factorize the covariance matrix. However, one could be interested in using more general covariance functions. In this section, we describe how the use of supervoxels can help to sample such Gaussian processes.

For instance, it can be of interest to sample segmentations whose smoothness is spatially inhomogeneous, when a structure presents both a regularly shaped outline and more irregular parts. To model this, the covariance function introduced by [Gibbs \(1998\)](#) and detailed in [Williams and Rasmussen \(2006\)](#) is of particular interest,

$$\forall \mathbf{a}, \mathbf{b} \in \Omega, \\ \Sigma(\mathbf{a}, \mathbf{b}) = \omega_0 \prod_{d=1}^D \left(\frac{2l_d(\mathbf{a})l_d(\mathbf{b})}{l_d^2(\mathbf{a}) + l_d^2(\mathbf{b})} \right)^{\frac{1}{2}} \exp \left(- \sum_{d=1}^D \frac{(\mathbf{a}_d - \mathbf{b}_d)^2}{l_d^2(\mathbf{a}) + l_d^2(\mathbf{b})} \right) \quad (13)$$

where D is the dimension of the input ($D = 2$ for an image and $D = 3$ for a volume), and l_d is any positive function defining a spatially varying correlation length. Note that if $l_d = p$ is a constant, we retrieve the squared exponential covariance function, $\Sigma(\mathbf{a}, \mathbf{b}) = \omega_0 \exp(-(\|\mathbf{a} - \mathbf{b}\|^2)/(2p^2))$.

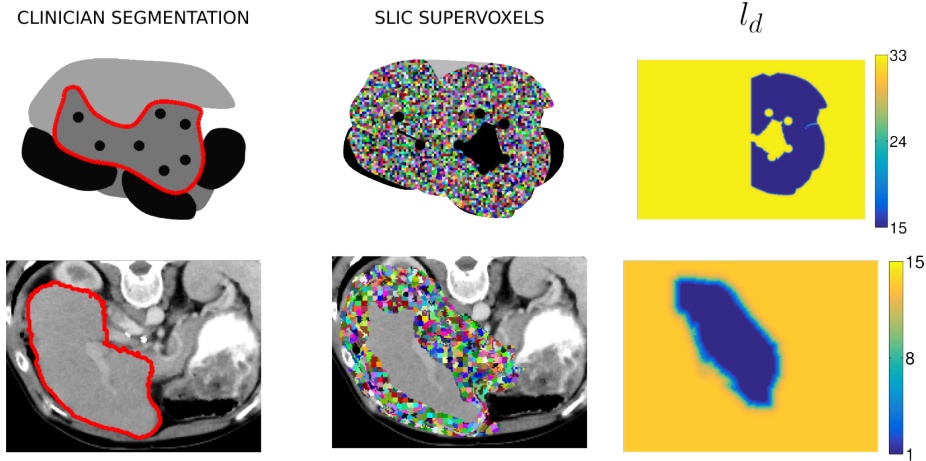


Figure 8: From left to right: clinician segmentation of a liver on a CT image, SLIC supervoxels used for the sampling, correlation length $l_1 = l_2 = l_3 = l_d$ for two different slices. For the liver, the correlation length is set to a lower value away from the rib cage to account for the noisier aspect of the contour.

However, for spatially varying l_d , the covariance function is neither stationary nor separable. To make the computations tractable, we propose to use supervoxels. The method is as follow:

- Compute the signed geodesic distance as before.
- Decompose the image into supervoxels having roughly homogeneous intensity. We use the SLIC algorithm developed by [Achanta et al. \(2012\)](#) which provides realistic supervoxels in linear time. We extended the algorithm such that the supervoxels respect the input ROI boundary. This is done by limiting the nearest neighbor search to voxels belonging to similar ROIs (ROI and background).
- Compute the covariance matrix defined on the set \mathcal{B} of barycenters of the supervoxels.
- Sample the Gaussian process defined over the set of supervoxels with covariance matrix Σ by computing the Cholesky decomposition of Σ , and interpolate it using a triangulation-based linear interpolation.

Figure 8 shows the clinician segmentation, the computed supervoxels, and the correlation length l_d for a 2D synthetic case, and one slice of a 3D liver CT scan. To speed up the computations, only the supervoxels around the boundary are considered, for a total of 4, 279 supervoxels for the synthetic image, and 14, 528 for the liver. The spatially varying correlation length l_d is set equal for every $d = 1, 2, 3$. A smaller correlation length has been set on a region of the liver which is expected to be less regular (away from the rib cage). Two examples of the resulting samples are presented on Figure 9. To emphasize the spatial difference in regularity, we also present the level set function variation around the mean $\varphi - \mu$ for the sample $\varphi \sim \mathcal{GP}(\mu, \Sigma)$.

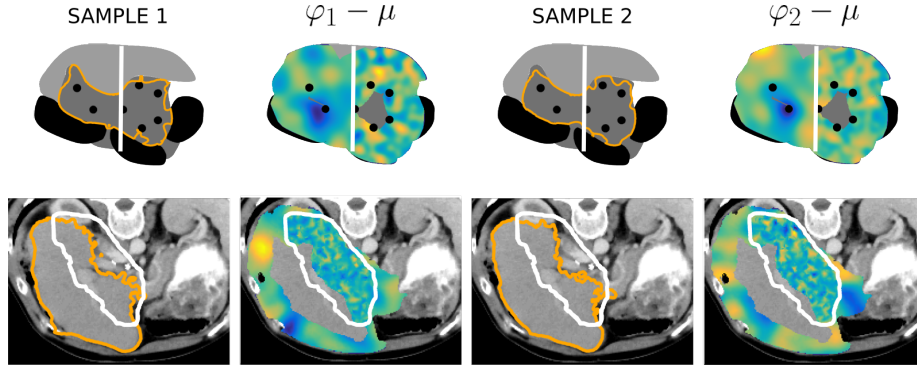


Figure 9: (Top) Synthetic case; (Bottom) 3D liver segmentation. From left to right: sample of a segmentation, corresponding level set variation $\varphi_1 - \mu$ with $\varphi_1 \sim \mathcal{GP}(\mu, \Sigma)$, additional sample, corresponding level set variation $\varphi_2 - \mu$ with $\varphi_2 \sim \mathcal{GP}(\mu, \Sigma)$. For clarity, the region of low correlation length is emphasized in white.

6.5. Plausibility of the Samples: Evaluation on multiple segmentations

In this section, we present an assessment of the visual plausibility of the samples, using tumor segmentations from the BraTS challenge ([Menze et al. \(2015\)](#)).

In other words, we test if the manual segmentations of the same structure could have been generated by GPSSI, if one of the segmentations was considered as the true one.

Algorithm 1: Sampling segmentations close to the target

Data: Source segmentation \mathcal{S}_0 ; target segmentation \mathcal{S} ; proposal distributions Q_1 , Q_2 , and Q_3 ; number of iterations K
Result: Samples close to the target segmentation
initialize ω_0 , ω_1 , and ϵ ;
while $r \leq K$ **do**
 $\hat{\omega}_0 = Q_1(\omega_0)$;
 $\hat{\omega}_1 = Q_2(\omega_1)$;
 $\hat{\epsilon}_0 = Q_3(\epsilon)$;
 Sample $\mathcal{S}^r \sim \mathcal{GP}(\mu(\mathcal{S}_0) + \hat{\epsilon}_0, \Sigma(\hat{\omega}_0, \hat{\omega}_1))$;
 Compute the likelihood $L^r = P(\mathcal{S}^r | \hat{\epsilon}_0, \hat{\omega}_0, \hat{\omega}_1)$;
 Compute the acceptance ratio $\mathcal{A} = \min(1, L^r / L^{r-1})$;
 Sample u uniform between 0 and 1 ;
 if $u < \mathcal{A}$ **then**
 $\omega_0 = \hat{\omega}_0$;
 $\omega_1 = \hat{\omega}_1$;
 $\epsilon = \hat{\epsilon}_0$;
 end
 $r+ = 1$;
end

Consider a tumor which has been manually segmented by $k + 1$ experts (Figure 10). Noting \mathcal{S}_i the segmentation from expert i , for $i = 0, \dots, k$, the goal is to generate segmentations close to $\{\mathcal{S}_i\}_{i=1, \dots, k}$ from the input \mathcal{S}_0 , assumed to be the reference segmentation. More precisely, we assume that

$$\text{for } i = 1, \dots, k \quad \mathcal{S}_i \sim \mathcal{GP}(\mu(\mathcal{S}_0) + \epsilon^i, \Sigma(\omega_0^i, \omega_1^i)) \quad (14)$$

where ω_0^i and ω_1^i are the parameters of the squared exponential covariance function, ϵ^i is the parameter taking into account over or under segmentation, and $\mu(\mathcal{S}_0)$ is the signed geodesic distance from \mathcal{S}_0 . For clarity of the notations, we drop the i index for the parameters in the rest of the paper. Since ω_0 , ω_1 , and ϵ are unknown, we resort to Markov Chain Monte Carlo (MCMC) sampling to produce samples close to the target segmentation \mathcal{S} . Following Bayes rule, and assuming independence of the parameters, we have

$$P(\epsilon, \omega_0, \omega_1 | \mathcal{S}) \propto P(\mathcal{S} | \epsilon, \omega_0, \omega_1) P(\epsilon) P(\omega_0) P(\omega_1) \quad (15)$$

We use uniform priors on $[-50, 50]$ for ϵ , $[1, 10000]$ for ω_0 , and $[1, 500]$ for ω_1 . We use normal proposal distributions Q_1 , Q_2 , and Q_3 for ϵ , ω_0 , and ω_1 respectively with a standard deviation of 2 for the three proposal distributions, set to reach an acceptance rate of approximately 50%. We define the likelihood,

$$P(\mathcal{S}_i|\epsilon, \omega_0, \omega_1) = \exp\left(-\frac{H(\epsilon, \omega_0, \omega_1)^2}{\sigma^2}\right) \quad (16)$$

where $H(\epsilon, \omega_0, \omega_1)$ is the 99th percentile of the symmetric Hausdorff distance between the boundary of \mathcal{S}_i and the boundary of a random sample from $\mathcal{GP}(\mu(\mathcal{S}_0) + \epsilon^i, \Sigma(\omega_0^i, \omega_1^i))$, and σ is a noise level set to 4 mm. Note that this likelihood is handcrafted because we observe contours of the clinician segmentations, and not level set functions. The algorithm is described in details in Algorithm 1.

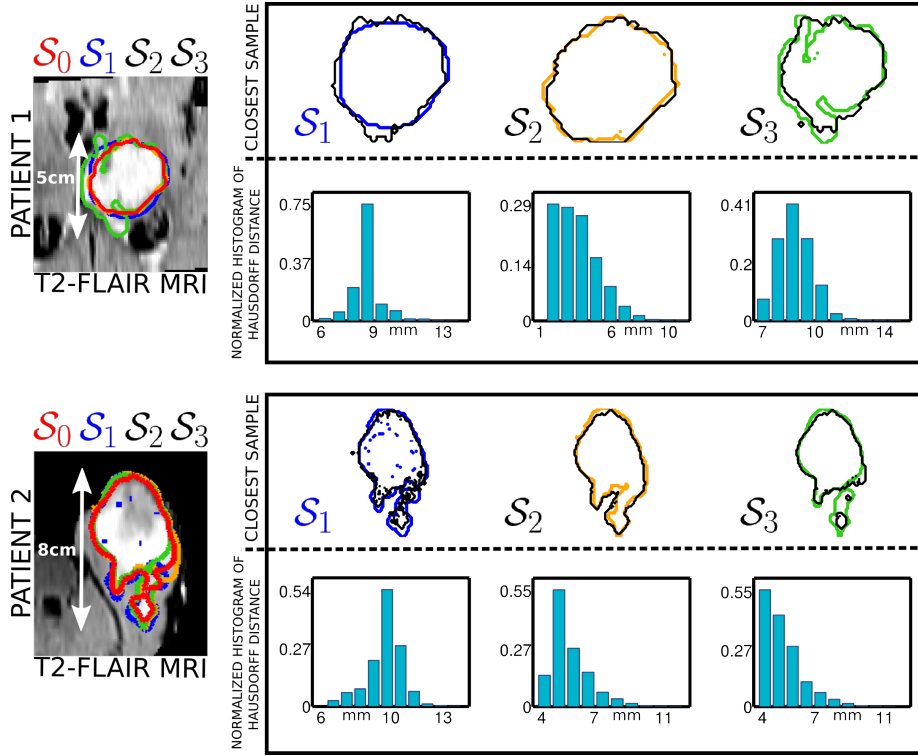


Figure 10: Patient 1 on top, Patient 2 on the bottom. From left to right: 4 clinician segmentations of a glioma on a T2-FLAIR MRI, the source segmentation \mathcal{S}_0 is the red one; closest sample in terms of Hausdorff distance in black with the corresponding target segmentation \mathcal{S}_i , $i = 1, 2, 3$, and normalized histogram of the Hausdorff distance between the target segmentation and the 20,000 samples.

Results are shown on Figure 10 for two glioma patients which were segmented by $k + 1 = 4$ clinicians using 20000 samples. We randomly choose one of the four segmentations as the source segmentation, and consider the three others as targets. Figure 10 shows the samples which result in the lowest Hausdorff distance (i.e. highest likelihood) among the sampled ones. The samples are close to the clinician segmentations: but the fit between the best sample and

segmentation \mathcal{S}_i depends on how far \mathcal{S}_i is from \mathcal{S}_0 . Based on the normalized histogram of Hausdorff distance, one can see for instance, that for patient 1 \mathcal{S}_3 is clearly further from \mathcal{S}_0 , whereas for patient 2 it is \mathcal{S}_1 . Note also that the parameter ω_1 (contour smoothness) is not well captured probably due to the nature of the Hausdorff distance. Although increasing the number of samples could further improve the match between the best sample and observed contours, this experiment shows that GPSSI can generate samples already close to real observed ones.

7. Tumor Delineation Uncertainty in Radiotherapy

The proposed method is applied to the uncertainty quantification of radiotherapy planning. The standard of care for grade IV gliomas (Figure 4) is the delivery of 60 Gray (Gy) to the Clinical Target Volume (CTV) which is defined as a 2-3cm extension of the Gross Tumor Volume (GTV) visible on a T1Gd MRI (Mason et al. (2007)).

The following algorithm is applied to the patient shown in Figure 11:

- Sample a GTV and a brainstem segmentation sample from the clinician segmentations visible on Figure 11 (top left).
- Compute the CTV from the GTV. In order to take into account the natural boundaries of the tumor progression and its privileged paths of progression, we compute the CTV using the tumor tail extrapolation algorithm developed by Konukoglu et al. (2010), which models the infiltration of glioma cells in the brain parenchyma on a single time point. It takes into account the fact that glioma cells preferably invade white matter over gray matter, and that the cerebrospinal fluid and falx cerebri are boundaries for the tumor progression. More precisely, we compute the tumor cell density infiltration as the tail extrapolation from the sampled GTV. It is based on a segmentation of the brain into white matter, gray matter and cerebrospinal fluid. The CTV is then defined as the largest volume enclosed by an isovalue of the tumor cell density, which is fully included in the 2cm isotropic extension of the GTV. Finally, the sampled brainstem is excluded from the CTV.
- Compute the prescribed dose for the CTV as 60 Gy targeted inside the CTV and 0 Gy elsewhere.

These steps were applied for 50 samples. Figure 11 shows the mean target dose and its standard deviation. Several strategies could be applied to take into account the uncertainty in the GTV delineation. Generally, radiotherapy planning has to find a compromise between delivering radiation to the tumor, and avoiding dose to radiosensitive tissues. Visualization of dose uncertainty may guide the physician in this process. For example, the radiation dose could be reduced in regions of high uncertainty if this allows for dose reductions in radiosensitive organs, and thereby reduces the risk of side effects substantially.

Technically, the standard deviation of the target dose could be used in the optimization of the radiation beams to weight differently voxels at the border of the CTV where the dose target is less certain. Moreover, it is important to visualize areas which represent tumor with near certainty and should be treated with the prescribed dose. In the long term, tumor segmentation samples could be used for radiotherapy planning based on models of tumor control probability (TCP).

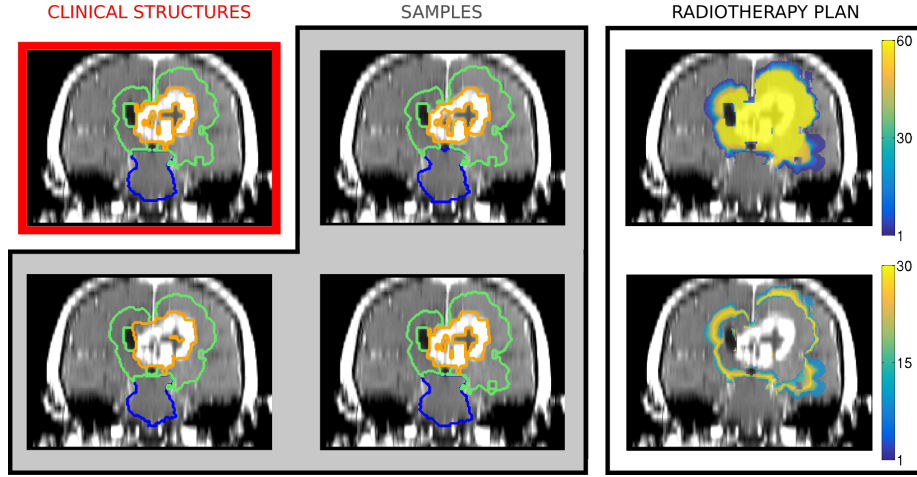


Figure 11: In the red box: the gross tumor volume (GTV) and the brainstem segmented by the clinician are in orange and blue respectively, the clinical target volume (CTV) is shown in green. In the gray box: samples of the GTV and brainstem in orange and blue respectively, corresponding CTV in green. In the white box, the average dose over 50 sampled GTV (top) and the dose standard deviation (bottom).

8. Conclusion

In this paper, an original image segmentation sampling framework has been proposed to generate plausible segmentations close to an input one. The approach leads to spatially smooth contours, which take into account the presence of salient features of the ROI in the image. Samples are efficiently generated, with a variability around a reference segmentation easily controlled by a single scalar. Several extensions have been proposed such as the sampling of several neighboring structures, taking into account under and over segmentation with a simple scalar parameter, and forcing the inclusion or the exclusion of certain part of the image from the samples. We provide a method to apply this sampling method to more general non-stationary covariance functions using supervoxels. The plausibility of the sampled contours originates from the inclusion of various constraints such as the intensity gradient, the presence of anatomical regions to be included or excluded, or the varying level of smoothness. To further improve the credibility of the samples, one could further restrict the choice of covariance functions based on observed multiple segmentations of the same structure as

hinted in section 6.5. Furthermore, this work could leverage the vast literature on Gaussian processes, imposing for instance convexity constraints on the samples (Da Veiga and Marrel (2012)). The application of the proposed method to segmentations based on multiple image modalities - such as combinations of MRI, CT, and PET - should also be investigated.

Future work will also explore the incorporation of uncertainty in the radiotherapy dose planning. The proposed method could have several additional applications for instance to produce consensus segmentations from several expert ones. It could also be used to assess the confidence of the performance of segmentation algorithms in the context of segmentation challenges, by providing several likely segmentations around the ground truth segmentations. Finally, this method could be included in model personalization pipeline (Lê et al. (2015a)) where segmentation is a crucial input. This would allow one to plan therapy based on personalized models which take into account the uncertainty in the data.

Acknowledgment

Part of this work was funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080. We would like to thank the organizers of the BraTS challenge for making the data publicly available, and Hannes Nickisch for his input regarding the sampling method during MICCAI 2015.

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34, 2274–2282.
- Chang, J., Fisher III, J.W., 2011. Efficient MCMC sampling with implicit shape representations, in: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, IEEE. pp. 2081–2088.
- Criminisi, A., Sharp, T., Blake, A., 2008. Geos: Geodesic image segmentation, in: *Computer Vision–ECCV 2008*. Springer, pp. 99–112.
- Da Veiga, S., Marrel, A., 2012. Gaussian process modeling with inequality constraints, in: *Annales de la Faculté des Sciences de Toulouse*, pp. 529–555.
- De Bruijne, M., Nielsen, M., 2004. Image segmentation by shape particle filtering, in: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, IEEE. pp. 722–725.
- Dietrich, C., Newsam, G.N., 1997. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing* 18, 1088–1107.

- Fan, A.C., Fisher III, J.W., Wells III, W.M., Levitt, J.J., Willsky, A.S., 2007. MCMC curve sampling for image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007. Springer, pp. 477–485.
- Gerardo-Castro, M.P., Peynot, T., Ramos, F., 2013. Laser-radar data fusion with Gaussian process implicit surfaces, in: Corke, P., Mejias, L., Roberts, J. (Eds.), The 9th International Conference on Field and Service Robotics, Brisbane, Australia.
- Gibbs, M.N., 1998. Bayesian Gaussian processes for regression and classification. Ph.D. thesis. University of Cambridge.
- Gilboa, E., Saatçi, Y., Cunningham, J.P., 2015. Scaling multidimensional inference for structured gaussian processes. Pattern Analysis and Machine Intelligence, IEEE Transactions on 37, 424–436.
- Konukoglu, E., Clatz, O., Bondiau, P.Y., Delingette, H., Ayache, N., 2010. Extrapolating glioma invasion margin in brain magnetic resonance images: Suggesting new irradiation margins. Medical image analysis 14, 111–125.
- Kozintsev, B., 1999. Computations with Gaussian random fields. Ph.D. thesis. University of Maryland.
- Kozintsev, B., Kedem, B., 2000. Generation of similar images from a given discrete image. Journal of Computational and Graphical Statistics 9, 286–302.
- Lê, M., Delingette, H., Kalpathy-Cramer, J., Gerstner, E.R., Batchelor, T., Unkelbach, J., Ayache, N., 2015a. Bayesian personalization of brain tumor growth model, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Springer, pp. 424–432.
- Lê, M., Unkelbach, J., Ayache, N., Delingette, H., 2015b. Gpssi: Gaussian process for sampling segmentations of images, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Springer, pp. 38–46.
- Lorenzi, M., Ziegler, G., Alexander, D.C., Ourselin, S., 2015. Efficient Gaussian process-based modelling and prediction of image time series, in: Information Processing in Medical Imaging, Springer. pp. 626–637.
- Mason, W., Del Maestro, R., Eisenstat, D., Forsyth, P., Fulton, D., Laperrière, N., Macdonald, D., Perry, J., Thiessen, B., Committee, C.G.R., et al., 2007. Canadian recommendations for the treatment of glioblastoma multiforme. Current Oncology 14, 110.
- McInerney, T., Terzopoulos, D., 1996. Deformable models in medical image analysis: a survey. Medical image analysis 1, 91–108.

- Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., et al., 2015. The multimodal brain tumor image segmentation benchmark (BRATS). *Medical Imaging, IEEE Transactions on* 34, 1993–2024.
- Niethammer, M., Pohl, K.M., Janoos, F., Wells III, W.M., 2015. Active mean fields for probabilistic image segmentation: Connections with chan-vese and rudin-osher-fatemi models. *arXiv preprint arXiv:1501.05680* .
- Petersen, K., Nielsen, M., Brandt, S.S., 2010. A static SMC sampler on shapes for the automated segmentation of aortic calcifications, in: *Computer Vision–ECCV 2010*. Springer, pp. 666–679.
- Pohl, K.M., Fisher, J., Bouix, S., Shenton, M., McCarley, R.W., Grimson, W.E.L., Kikinis, R., Wells, W.M., 2007. Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis* 11, 465–477.
- Saatçi, Y., 2012. Scalable inference for structured Gaussian process models. Ph.D. thesis. University of Cambridge.
- Sabuncu, M.R., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. *Medical Imaging, IEEE Transactions on* 29, 1714–1729.
- Toivanen, P.J., 1996. New geodesic distance transforms for gray-scale images. *Pattern Recognition Letters* 17, 437–450.
- Vese, L.A., Chan, T.F., 2002. A multiphase level set framework for image segmentation using the mumford and shah model. *International journal of computer vision* 50, 271–293.
- Wang, C., Komodakis, N., Paragios, N., 2013. Markov random field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding* 117, 1610–1627.
- Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on* 23, 903–921.
- Williams, C.K., Rasmussen, C.E., 2006. Gaussian processes for machine learning. Cambridge: MIT Press .
- Williams, O., Fitzgibbon, A., 2007. Gaussian process implicit surfaces, in: *Gaussian Process in Practice*.
- Zhao, H.K., Chan, T., Merriman, B., Osher, S., 1996. A variational level set approach to multiphase motion. *Journal of computational physics* 127, 179–195.